Software That Supports Novel Integration, Unique Querying Ability, and Analysis of Locally Held Genetic, Experimental, and Clinical Data

Lamers, Susanna L.*, Dunlap L., Compton R., Beason S.
Gene Genie Incorporated, Thibodaux, LA, USA

Research of the Human Immunodeficiency Virus (HIV) involves ongoing, repetitious sequencing of the HIV genome and the mass accumulation of associated investigational data. At clinical research laboratories, this has frequently led to large amounts of viral information in varied formats, disordered data collection, and a desire for new data storage and analysis methods. Our objective was to develop tools specific for HIV research that focus on genetic data management and analysis. Our result, software designated HIVbase®, uses familiar, easy-to-use interfaces and provides applications for integrating HIV genetic, experimental, and clinical data. HIVbase contains components that allow for the mass import of raw genetic data, tools that eliminate repetitious sequence editing, tools that have the ability to automatically identify and store HIV regions of interest from nucleic acid or protein sequences, and tools that allow for the export of data in commonly used, analysis-ready formats. Storing data in a relational format allowed for further development of unique querying approaches. Our data base querying tool allows users to formulate novel questions that can be applied towards thousands of internally held sequences, to retrieve large amounts of genetic data based on specific characteristics, and to transform unwieldy raw data into formats that facilitate diagnosis and discovery. The database architecture uses a well-established business model that incorporates distributed functionality: 1) a Data Access Layer, 2) a Business Logic Layer, and 3) a User interface. These components contain the major blocks of software functionality needed to fulfill the many requests made by the user and the data model is robust enough to handle the multitude of data types that can be associated with sequence data**.** The database was built using Microsoft SQL Server 2000 Desktop Engine and related technologies. Both the Business Layer and the User Interface were constructed using Microsoft's .NET framework. These programs offer high-speed, diverse functionality and state-of-the-art software design. We now have the ability to expand upon this software and incorporate new analysis algorithms that can be applied to the data held within.

Grant Support: National Science Foundation, SBIR grant 0232127; National Institutes of Health, SBIR grant R43-AI-56904.